CS 726: Homework #4

Posted: 03/08/2020, due: 03/27/2020 by 5pm on Canvas

Please typeset or write your solutions neatly! If we cannot read it, we cannot grade it.

Note: You can use the results we have proved in class – no need to prove them again.

Q 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be an L-smooth function and let $\mathcal{X} \subseteq \mathbb{R}^n$ be a closed, convex, and nonempty set. Recall the definition of the gradient mapping: $G_\eta(\mathbf{x}) = \eta(\mathbf{x} - P_{\mathcal{X}}(\mathbf{x} - \frac{1}{\eta}\nabla f(\mathbf{x})))$, where $P_{\mathcal{X}}(\cdot)$ denotes the Euclidean projection onto $\mathcal{X}$. Prove that $G_\eta(\cdot)$ is $(2\eta + L)$-Lipschitz continuous. [20pts]

Q 2. Consider a constrained minimization problem $\min_{\mathbf{u}\in\mathcal{X}} f(\mathbf{x})$, where $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth and $\mathcal{X} \subseteq \mathbb{R}^n$ is a hyper-rectangle: $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : x_i \in [a_i, b_i], \forall i \in \{1, 2, \dots, n\}\}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are vectors that satisfy $\forall i \in \{1, \dots, n\} : a_i < b_i$.

(i) What is $P_{\mathcal{X}}(\mathbf{x})$? Write it in closed form. [10pts]

(ii) Define $\Delta_i(\mathbf{x}) := \nabla_i f(\mathbf{x})\mathbf{e}_i$, where $\mathbf{e}_i$ is the i^{th} standard basis vector (except for its i^{th} coordinate, all coordinates are equal to zero; the i^{th} coordinate equals one). Define:

$$T(\mathbf{x}, i) := \operatorname*{argmin}_{\mathbf{u}\in\mathcal{X}} \Big\{ \langle \Delta_i(\mathbf{x}), \mathbf{u} - \mathbf{x}\rangle + \frac{L}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 \Big\}.$$

Express the gradient mapping $G_L(\mathbf{x})$ as a function of L, $\mathbf{x}$, and $T(\mathbf{x}, i)$, $i \in \{1, \dots, n\}$. [10pts]

(iii) Consider the following method that starts from some initial point $\mathbf{x}_0 \in \mathcal{X}$ and updates its iterates $\mathbf{x}_k$ for $k \geq 0$ as:

$$i_k^* = \operatorname*{argmax}_{1\leq i\leq n} |(G_L(\mathbf{x}_k))_i|$$
$$\mathbf{x}_{k+1} = T(\mathbf{x}_k, i_k^*).$$

Prove the following sufficient descent property for this algorithm:

$$(\exists \alpha > 0)(\forall k \geq 0) : \quad f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{\alpha}{2}\|G_L(\mathbf{x}_k)\|_2^2.$$

What is the largest α for which this property holds? What can you say about convergence of this method if f is bounded below by some $\bar{f}$? [10pts]

Q 3. Consider a constrained minimization problem $\min_{\mathbf{u}\in\mathcal{X}} f(\mathbf{x})$, where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and M-Lipschitz, and $\mathcal{X} \subseteq \mathbb{R}^n$ is closed, convex, and nonempty. Assume that there exists $\mathbf{x}^* \in \mathcal{X}$ that minimizes f. Recall from the class that we use $\mathbf{g}_{\mathbf{x}}$ to denote an arbitrary subgradient of f at $\mathbf{x} \in \mathcal{X}$. Distances within the set $\mathcal{X}$ are measured using some norm $\|\cdot\|$ (e.g., an ℓ_p norm for $p \geq 1$). Norm that is dual to $\|\cdot\|$ is denoted by $\|\cdot\|_*$ and is defined by: $\|\mathbf{z}\|_* = \max_{\mathbf{u}:\|\mathbf{u}\|\leq 1} \langle \mathbf{z}, \mathbf{u}\rangle$. By the definition of a dual norm, you can immediately deduce that:

$$(\forall \mathbf{u}, \mathbf{z}) : \langle \mathbf{u}, \mathbf{z}\rangle \leq \|\mathbf{u}\|\|\mathbf{z}\|_*.$$

Note also that since f is M-Lipschitz w.r.t. the norm $\|\cdot\|$, we have, $\forall \mathbf{x}, \mathbf{y}$ and all $\mathbf{g}_{\mathbf{x}} \in \partial f(\mathbf{x})$:

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq M\|\mathbf{x} - \mathbf{y}\| \quad \text{ and } \|\mathbf{g}_{\mathbf{x}}\|_* \leq M.$$

In this part, you will analyze the convergence of the Mirror Descent algorithm, defined by:

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{u}\in\mathcal{X}} \{a_k \langle \mathbf{g}_{\mathbf{x}_k}, \mathbf{u} - \mathbf{x}_k\rangle + D_\psi(\mathbf{u}, \mathbf{x}_k)\}, \tag{MD}$$

where, as in previous assignments, $D_\psi(\mathbf{x}, \mathbf{y})$ denotes the Bregman divergence between $\mathbf{x}$ and $\mathbf{y}$ w.r.t. a continuously-differentiable function ψ.

Assume that ψ is 1-strongly convex w.r.t. $\|\cdot\|$, namely:

$$(\forall \mathbf{x}, \mathbf{y}): \quad \psi(\mathbf{y}) \geq \psi(\mathbf{x}) + \langle \nabla\psi(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

(i) Using the first-order optimality in the definition of $\mathbf{x}_{k+1}$ and the three-point identity of Bregman divergences we proved in HW #2, show that:

$$(\forall \mathbf{u} \in \mathcal{X}): \quad a_k \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{u}\rangle \leq D_\psi(\mathbf{x}^*, \mathbf{x}_k) - D_\psi(\mathbf{x}^*, \mathbf{x}_{k+1}) - D_\psi(\mathbf{x}_{k+1}, \mathbf{x}_k). \qquad \text{[15pts]}$$

(ii) Use Part (i) to prove that:

$$a_k \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}^*\rangle \leq \frac{{a_k}^2 M^2}{2} + D_\psi(\mathbf{x}^*, \mathbf{x}_k) - D_\psi(\mathbf{x}^*, \mathbf{x}_{k+1}). \qquad \text{[15pts]}$$

(iii) Use Part (ii) to prove that, $\forall k \geq 0:$

$$f(\mathbf{x}_k^{\mathrm{out}}) - f(\mathbf{x}^*) \leq \frac{D_\psi(\mathbf{x}^*, \mathbf{x}_0) + \frac{M^2}{2}\sum_{i=0}^k {a_i}^2}{A_k},$$

where $\mathbf{x}_k^{\mathrm{out}} = \frac{\sum_{i=0}^k a_i \mathbf{x}_i}{A_k}$. [10pts]

(iv) Discuss how you would choose $\{a_i\}_{i=0}^k$ for (MD) to converge as fast as possible, and provide the convergence bound (a bound on $f(\mathbf{x}_k^{\mathrm{out}}) - f(\mathbf{x}^*)$) for your choice(s) of $\{a_i\}_{i=0}^k$. [10pts]